

Università degli Studi di Napoli Federico II  
Scuola delle Scienze Umane e Sociali  
Quaderni  
11

**ASMOD 2018**  
**Proceedings of the International Conference on  
Advances in Statistical Modelling of Ordinal Data**  
Naples, 24-26 October 2018

Editors  
Stefania Capecchi, Francesca Di Iorio, Rosaria Simone



Federico II University Press







Università degli Studi di Napoli Federico II  
Scuola delle Scienze Umane e Sociali  
Quaderni

11



ASMODO 2018  
Proceedings of the International Conference on  
Advances in Statistical Modelling of Ordinal Data  
Naples, 24-26 October 2018

Editors

Stefania Capecchi, Francesca Di Iorio, Rosaria Simone

Federico II University Press



fedOA Press

ASMOD 2018 : Proceedings of the Advanced Statistical Modelling for Ordinal Data Conference : Naples, 24-26 October 2018 / editors Stefania Capecchi, Francesca Di Iorio, Rosaria Simone. – Napoli : FedOAPress, 2018. – (Scuola di Scienze Umane e Sociali. Quaderni ; 11).

Accesso alla versione elettronica:

<http://www.fedoabooks.unina.it>

ISBN: 978-88-6887-042-3

DOI: 10.6093/978-88-6887-042-3

ISSN Collana: 2499-4774

#### Comitato scientifico

Enrica Amatore (Università di Napoli Federico II), Simona Balbi (Università di Napoli Federico II), Antonio Blandini (Università di Napoli Federico II), Alessandra Bulgarelli (Università di Napoli Federico II), Adele Caldarelli (Università di Napoli Federico II), Aurelio Cernigliaro (Università di Napoli Federico II), Lucio De Giovanni (Università di Napoli Federico II), Roberto Delle Donne (Università di Napoli Federico II), Arturo De Vivo (Università di Napoli Federico II), Oliver Janz (Freie Universität, Berlin), Tullio Jappelli (Università di Napoli Federico II), Paola Moreno (Université de Liège), Edoardo Massimilla (Università di Napoli Federico II), José González Monteagudo (Universidad de Sevilla), Enrica Morlicchio (Università di Napoli Federico II), Marco Musella (Università di Napoli Federico II), Gianfranco Pecchinenda (Università di Napoli Federico II), Maria Laura Pesce (Università di Napoli Federico II), Domenico Piccolo (Università di Napoli Federico II), Mario Rusciano (Università di Napoli Federico II), Mauro Sciarelli (Università di Napoli Federico II), Roberto Serpieri (Università di Napoli Federico II), Christopher Smith (British School at Rome), Francesca Stroffolini (Università di Napoli Federico II), Giuseppe Tesauro (Corte Costituzionale)

© 2018 FedOAPress - Federico II Open Access University Press

Università degli Studi di Napoli Federico II  
Centro di Ateneo per le Biblioteche “Roberto Pettorino”  
Piazza Bellini 59-60  
80138 Napoli, Italy  
<http://www.fedoapress.unina.it/>

Published in Italy

Gli E-Book di FedOAPress sono pubblicati con licenza  
Creative Commons Attribution 4.0 International

## Contents

Foreword	ix
<i>Invited papers</i>	
A. Agresti, <i>Simple ordinal model effect measures</i>	1
B. Francis, <i>Latent class models for multiple ordinal items</i>	3
B. Grün, G. Malsiner-Walli, <i>Bayesian latent class analysis with shrinkage priors: an application to the Hungarian heart disease data</i>	13
M. Kateri, <i>Modelling Ordinal Data: A <math>\phi</math>-divergence based approach</i>	25
E. Ronchetti, <i>Robust statistical analysis of ordinal data</i>	27
G. Tutz, <i>Uncertainty, dispersion and response styles in ordinal regression</i>	33
<i>Accepted papers</i>	
A. Barbiero, <i>Inducing a desired value of correlation between two point-scale variables</i>	45
A. Bonanomi, M.N. Ruscone, S.A. Osmetti, <i>Dissimilarity measure for ranking data via mixture of copulae</i>	53
G. Bove, E. Nuzzo, A. Serafini, <i>Measurement of interrater agreement for the assessment of language proficiency</i>	61

E. Brentari, M. Manisera, P. Zuccolotto, <i>Modelling perceived variety in a choice process with nonlinear CUB</i>	69
R. Colombi, S. Giordano, <i>A flexible distribution to handle response styles when modelling rating scale data</i>	77
M. Corduas, <i>Joint modelling of ordinal data: a copula-based method</i>	85
C. Davino, T. Naes, R. Romano, D. Vistocco, <i>Modeling preferences: beyond the average effects</i>	93
C. Davino, R. Simone, D. Vistocco, <i>Exploring synergy between CUB models and quantile regression: a comparative analysis through continuousized data</i>	101
E. di Bella, L. Leporatti, F. Maggino, L. Gandullia, <i>A Poset based indicator of gender equality at sub-national level</i>	109
M. Fattore, A. Arcagni, <i>Using mutual ranking probabilities for dimensionality reduction and ranking extraction in multidimensional systems of ordinal variables</i>	117
S. Golia, M. Carpita, <i>On classifiers to predict soccer match results</i>	125
L. Grilli, M.F. Marino, O. Paccagnella, C. Rampichini, <i>Multiple imputation and selection of ordinal level-2 predictors in multilevel models</i>	133
M. Iannario, A.C. Monti, P. Scalera, <i>Why the number of response categories in rating scales should be large</i>	139

A. Lorenzo-Arribas, M.J. Brewer, A.M. Overstall, <i>Solutions to issues with partial proportional odds models</i>	147
F. Pennoni, M. Nakai, <i>A latent variable model for a derived ordinal response accounting for sampling weights, missing values and covariates</i>	155
F. Pesarin, L. Salmaso, H. Huang, R. Arboretti, R. Ceccato, <i>Permutation tests for stochastic ordering with ordinal data</i>	163
A. Plaia, M. Sciandra, S. Buscemi, <i>Consensus measures for preference rankings with ties: an approach based on position weighted Kemeny distance</i>	171
M. Ranalli, R. Rocci, <i>Simultaneous clustering and dimensional reduction of mixed-type data</i>	179
V. Sansivieri, M. Matteucci, S. Mignani, <i>Bi-Factor MIRT observed-score equating under the NEAT design for tests with several content areas</i>	187
R. Simone, F. Di Iorio, C. Cappelli, <i>On the choice of splitting rules for model-based trees for ordinal responses</i>	195
F. Torti, S. Salini, M. Riani, <i>Robustness issues for categorical data</i>	203
M. Vives-Mestres, J. Antoni Martín-Fernández, S. Thió-Henestrosa, R.S. Kenett, <i>Applications and theoretical results of association rules and compositional data analysis: a contingency table perspective</i>	211



# A flexible distribution to handle response styles when modelling rating scale data

Roberto Colombi\*, Sabrina Giordano\*\*

*Abstract:* It is commonly known that the respondents to rating scale questions, when are not aware, can select their own response using only certain response categories regardless the item content. This behavior is described as response style. Thus, the observed response can be a real opinion or dictated by a response style behavior. Marginal models (HMMLU) for multivariate responses by Colombi et al., 2018, enables us to distinguish these two behaviors and allows to specify the distributions of uncertain responses. We extend the class of HMMLU models with a new family of discrete distributions whose two parameters allow the uncertain distributions to be U-shaped, bell-shaped, unimodal, symmetric, skewed or uniform, for capturing different response styles.

*Keywords:* Mixture models, Latent variables, Marginal models.

## 1. Introduction

Questionnaires with rating scale items are widely used in psychological, social or marketing surveys to measure opinions, interests, or attitudes. In such contexts, it is commonly observed that a respondent, when in doubt, may consistently use only a few of the given options irrespective of his/her opinion. Someone may skip the endpoints, others have tendency to mark the extremes or the middle category (extreme or midpoint response styles), others respond with agreement/disagreement (acquiescence) regardless of item content, optimists may overvalue their feelings and pessimists may underrate them (one side contraction). The term response style indicates this systematic tendency and it is extensively debated in the literature (e.g. Baumgartner and Steenkamp, 2001).

A family of marginal models (HMMLU) for multivariate responses has been introduced by Colombi et al., 2018, to take into account that an ob-

\*University of Bergamo, colombi@unibg.it

\*\*University of Calabria, sabrina.giordano@unical.it

served response can be the real respondent’s attitude (aware response) or ensued from a response style. An HMMLU model enables us to specify the distribution of responses due to response styles (called uncertainty distribution) and distinguish it from that of responses dictated by awareness. Uniform and shifted Parabolic probability functions (Colombi et al., 2018) have been used as uncertainty distributions in the HMMLU model. Since the proposed distributions are symmetric or can cope with only one response style, the class of HMMLU models is, in this paper, enriched by a new family of distributions with two shape parameters. This gives the opportunity of choosing among several alternatives of uncertainty distributions (U-shaped, bell-shaped, unimodal, symmetric, skewed, uniform distributions) which capture different response styles. Covariates are also inserted to account for individual differences in response styles.

2. *The family of shifted reshaped parabolic distributions*

Remind that a probability function  $p(i)$ ,  $i = 1, 2, \dots, m$ , of a discrete variable with  $m$  levels can be specified by a set of local logits  $l_i$ ,  $i = 1, 2, \dots, m - 1$ , as shown below

$$p(1) = \frac{1}{1 + \sum_{j=2}^m \exp\{\sum_{i=1}^{j-1} l_i\}}, p(i) = \frac{\exp\{\sum_{j=1}^{i-1} l_j\}}{1 + \sum_{j=2}^m \exp\{\sum_{i=1}^{j-1} l_i\}}, i = 2, \dots, m.$$

We derive a new family of distributions by a linear transformation of the local logits of the Parabolic random variable with probability function

$$p(i) = \frac{6(m + 1 - i)i}{(m + 2)(m + 1)m}, \quad i = 1, 2, \dots, m.$$

More precisely the Local Shifted Reshaped Parabolic (LSRP) distribution is specified by the local logits  $l_i$  given by the linear transformation

$$l_i = \phi_0 + \phi_1 \log \frac{p(i + 1)}{p(i)}, \quad i = 1, 2, \dots, m - 1.$$

The LSRP distribution family contains, as a special case, the Uniform distribution ( $\phi_0 = \phi_1 = 0$ ), while for negative (positive) values of  $\phi_1$  it is U-

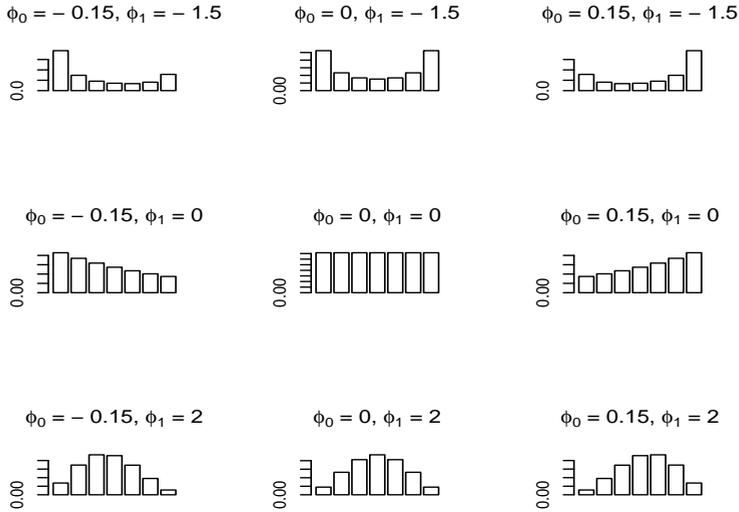


Figure 1. Local Shifted Reshaped Parabolic distributions with different shape parameters

shaped (bell-shaped).

Thus, parameter  $\phi_1$  rules the frequencies for extreme and middle points. Specifically, high (low) values of  $\phi_1$  correspond to distributions where indecision leads to focus on middle categories (extreme categories). This allows us to describe adequately extreme and midpoints response styles. Parameter  $\phi_0$  governs the skewness of the LSRP distributions. In fact, if  $\phi_0 = 0$ , these distributions are symmetric, right skewed for  $\phi_0 > 0$  and left skewed otherwise. Moreover, for a given  $\phi_1$ , LSRP distributions are monotonically ordered, according to the likelihood ratio stochastic ordering as function of  $\phi_0$ . Positive (negative) values enable to capture the acquiescence response style of respondents who tend to endorse the agreement (disagreement) side of the rating scale. Figure 1 shows some examples. Following the same reasoning, distributions like LSRP can be obtained starting from other probability functions independent from unknown parameters or using different logits. The idea of deriving a new distribution by transforming linearly logits of a free parameter discrete probability function is convenient when marginal models are used to

fit ordinal data. Alternative uncertainty distributions (e.g. Tutz and Schneider, 2017) could be proposed in the context of HMMLU models but with less advantages in terms of computational ease.

### 3. A mixture model with LSRP uncertainty distributions

We present the simple bivariate version of the HMMLU model, introduced by Colombi et al., 2018, in a more general extent.

Let  $R_1$  and  $R_2$  be two ordinal variables with support  $\{1, 2, \dots, m_1\}$  and  $\{1, 2, \dots, m_2\}$ , respectively. We assume the existence of two binary latent variables,  $U_l$ ,  $l = 1, 2$ , such that the respondent answers the  $l^{th}$  question according to his/her awareness when  $U_l = 1$  or his/her response style when  $U_l = 0$ . We assume that each observable variable  $R_l$  depends only on its latent variable  $U_l$ ,  $l = 1, 2$ , and that the observable responses  $R_1$  and  $R_2$  are independent when at least one of them is given under uncertainty. Therefore, the joint distribution of the observable variables is specified by the mixture

$$\begin{aligned}
 P(R_1 = r_1, R_2 = r_2) &= \pi_{00} g_1(r_1, \phi_{01}, \phi_{11}) g_2(r_2, \phi_{02}, \phi_{12}) \\
 &+ \pi_{01} g_1(r_1, \phi_{01}, \phi_{11}) P(R_2 = r_2 \mid U_2 = 1) \\
 &+ \pi_{10} P(R_1 = r_1 \mid U_1 = 1) g_2(r_2, \phi_{02}, \phi_{12}) \\
 &+ \pi_{11} P(R_1 = r_1, R_2 = r_2 \mid U_1 = 1, U_2 = 1)
 \end{aligned} \tag{1}$$

for every  $r_1 = 1, 2, \dots, m_1$  and  $r_2 = 1, 2, \dots, m_2$ , where  $\pi_{ij} = P(U_1 = i, U_2 = j)$ ,  $i = 0, 1$ ,  $j = 0, 1$ , are the joint probabilities of the latent variables. Specifically they are the probabilities that both the answers are given with awareness ( $\pi_{11}$ ), both with uncertainty ( $\pi_{00}$ ) or one with uncertainty and the other one with awareness ( $\pi_{01}$  and  $\pi_{10}$ ). Moreover,  $g_l(r_l, \phi_{0l}, \phi_{1l})$ ,  $l = 1, 2$ , denotes the distribution of responses under uncertainty, which here belong to the family of Shifted Reshaped Parabolic introduced in Section 2. Finally,  $P(R_1 = r_1, R_2 = r_2 \mid U_1 = 1, U_2 = 1)$  is the joint distribution of the two aware responses and  $P(R_l = r_l \mid U_l = 1)$  are the marginal ones, with  $r_l = 1, 2, \dots, m_l$ ,  $l = 1, 2$ . The probabilities  $\pi_{ij}$ ,  $i = 0, 1$ ,  $j = 0, 1$ , are parameterized through two marginal logits  $\lambda_l$ ,  $l = 1, 2$ , measuring the probability of being uncertain on each specific item, plus a log odds ratio  $\lambda_{12}$ . When this parameter is positive, respondents tend to have the same behavior

of uncertainty/awareness on the two items. A marginal parametrization is also adopted for the joint probabilities  $P(R_1 = r_1, R_2 = r_2 \mid U_1 = 1, U_2 = 1)$ . So that, to parameterize the probabilities  $P(R_1 = r_1 \mid U_1 = 1)$ , and the probabilities  $P(R_2 = r_2 \mid U_2 = 1)$ , we introduce the vectors  $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2$  of  $(m_1 - 1)$  and  $(m_2 - 1)$  marginal local logits, respectively. These logits and the vector  $\boldsymbol{\eta}_{12}$  of  $(m_1 - 1)(m_2 - 1)$  local log odds ratios parameterize the joint distribution  $P(R_1 = r_1, R_2 = r_2 \mid U_1 = 1, U_2 = 1)$ . As the number of parameters is  $m_1 m_2 + 7$  the mixture is not identifiable without further constraints. If a set of covariates accounts for respondents heterogeneity, identifiability can be assured by linear models for the logits  $\lambda_l, l = 1, 2$  and the shape parameters  $\phi_{11}, \phi_{12}$ , parallel linear models for  $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2$  and by the assumption that  $\phi_{01}, \phi_{02}, \lambda_{12}$  do not depend on covariates (see Colombi et al., 2018, for more details). Useful restrictions on  $\boldsymbol{\eta}_{12}$  are the conditions of homogeneous and uniform association.

#### 4. An example

We analyse the data from the module on health and care seeking of the European Social Survey (ESS2 2004). Respondents are asked to reply to questions on alternative forms of health care, such as  $R_1 = \textit{Sex}$  (Approve if healthy people use medicines to improve sex life), and  $R_2 = \textit{Happy}$  (Approve if healthy people use medicines to feel happier). The responses are given on a 5-points scale (1 = “strongly approve”, 2 = “approve”, 3 = “Neither approve nor disapprove”, 4 = “disapprove”, 5 = “strongly disapprove”). In addition, we consider two explanatory variables: *Gender* (0 = “Male”, 1 = “Female”) and *Country* (0 = “France” and 1 = “United Kindom”).

We believe that the observed responses could be contaminated by some response styles. Thus, some HMMLU models are adapted to the data at hand in order to account for such behavior in the responses. We focus our attention also on detecting whether the shape of the uncertainty distributions varies according to the respondent’s characteristics.

With this aim, we fit HMMLU models specified under different hypotheses on  $g_l(r_l, \phi_{0l}, \phi_{1l}), l = 1, 2$  in model (1). The parameters are now denoted as  $\phi_{0l}^{GC}, \phi_{1l}^{GC}$ , with  $G, C = 0, 1, l = 1, 2$  in the strata identified by the combina-

tions of *Gender* and *Country*. We consider the following uncertainty distributions, *U*: Uniform ( $\phi_{0l}^{GC} = 0, \phi_{1l}^{GC} = 0, G, C = 0, 1, l = 1, 2$ ); *RP*: Reshaped Parabolic ( $\phi_{0l}^{GC} = 0, \phi_{1l}^{GC} = \phi_{1l}, G, C = 0, 1, l = 1, 2$ ); *SRP*: Shifted Reshaped Parabolic ( $\phi_{0l}^{GC} = \phi_{0l}, \phi_{1l}^{GC} = \phi_{1l}, G, C = 0, 1, G, C = 0, 1, l = 1, 2$ ); *HRP*: Heterogeneous Reshaped Parabolic ( $\phi_{0l}^{GC} = 0, \phi_{1l}^{GC} = \beta_{0l} + \beta_l^G + \beta_l^C, G, C = 0, 1, l = 1, 2$ ); *HSRP*: Heterogeneous Shifted Reshaped Parabolic ( $\phi_{0l}^{GC} = \phi_{0l}, \phi_{1l}^{GC} = \beta_{0l} + \beta_l^G + \beta_l^C, G, C = 0, 1, l = 1, 2$ ). These models are compared in Table 1, where it is evident that model  $\mathcal{M}_5$  shows the best fit.

Table 1. Models Comparison

Model	Unc. distr.	loglik	n.par.	Compared Models	LRT	p-value
$\mathcal{M}_1$	<i>U</i>	-9364.435	20			
$\mathcal{M}_2$	<i>RP</i>	-9198.537	22	$\mathcal{M}_1$ vs $\mathcal{M}_2$	331.795	0.0000
$\mathcal{M}_3$	<i>SRP</i>	-9171.389	24	$\mathcal{M}_2$ vs $\mathcal{M}_3$	54.2966	0.0000
$\mathcal{M}_4$	<i>HRP</i>	-9172.123	26	$\mathcal{M}_2$ vs $\mathcal{M}_4$	52.8283	0.0000
$\mathcal{M}_5$	<i>HSRP</i>	-9155.784	28	$\mathcal{M}_4$ vs $\mathcal{M}_5$	32.6785	0.0000

In each model the probabilities to give aware answers or to have a tendency towards a response style vary according to *Gender* and *Country* since there is parallel additive effect of the covariates on the logits mentioned in Section 3. The association among aware responses is assumed uniform homogeneous.

According to model  $\mathcal{M}_5$  respondents belonging to the four covariate groups behave differently when uncertain. Plots in Figure 2 illustrates the uncertainty distributions for the two items and the four strata. English respondents totally avoid the extremes and take a shelter in the neutral category when the question concerns the admissible use of pills to improve sex performances. In France, instead, for the same question people seem less elusive and concentrate their answers on the positive/intermediate side. As the matter refers happiness, unaware people, both in UK and France, distribute their preferences quite equally from “approve” to “disapprove”. In UK, the distribution is a bit more picked on the middle category. Extremes are not totally excluded, especially in France. Men and women show a very similar uncertain behavior in answering both the questions.

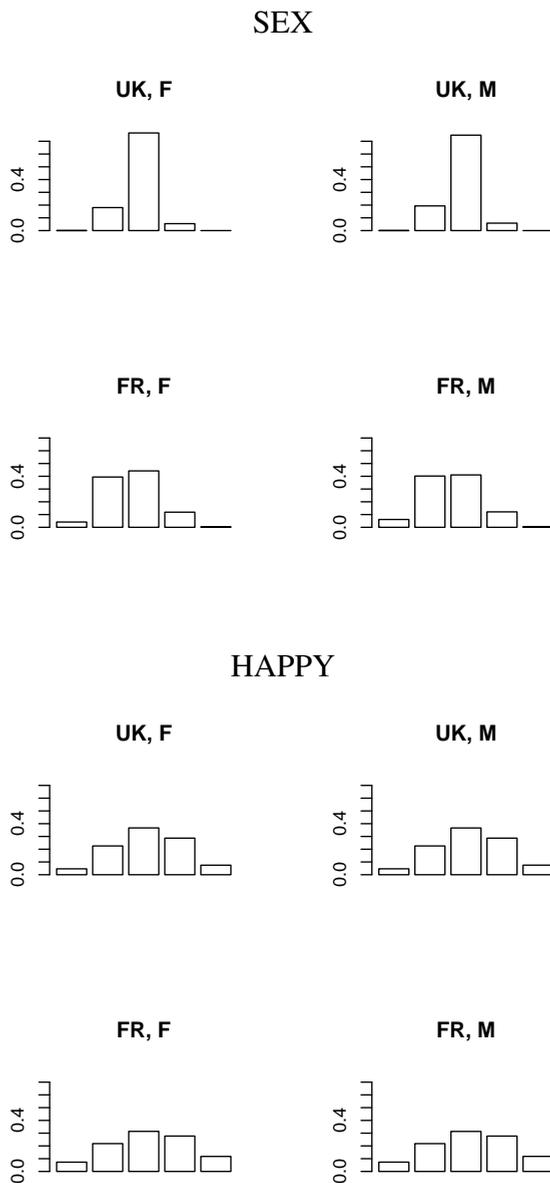


Figure 2. Uncertainty distributions of respondents belonging to the four Gender-Country groups for the two items: Sex and Happy

## References

- Baumgartner H., Steenkamp J.B. (2001) Response styles in marketing research: A cross-national investigation, *Journal of Marketing Research*, 38, 143-156.
- Colombi R., Giordano S., Gottard A., Iannario M. (2018) Hierarchical marginal models with latent uncertainty, [arxiv.org/pdf/1607.00882](https://arxiv.org/pdf/1607.00882).
- Tutz G., Schneider M. (2017) Mixture models for ordinal responses with a flexible uncertainty component, Technical Report Number 203, University of Munich.